LIST OF TABLES

6. High Level Milestones, Work Breakdown Structure and Budget for the User Facilities Subproject

The Tier 1 Regional Center will be managed within the Fermilab Computing Division, and falls under the responsibility of the Level 2 Project Manager for the User Facilities Subproject. The necessary R&D, prototyping, establishment and operations of the Regional Center are all part of the WBS for the US Software and Computing Project, which is described below. The Regional Center needs to be fully responsive to the needs of its constituencies. Sections 3-5 of this plan describe the points of contact with the US CMS and International CMS collaborations.

6.1 User Facilities High Level Milestones

A timeline for the establishment of the Tier 1 Regional Center, together with some relevant CMS milestones, are listed below:

Year	**USCMS activity** (by US fiscal year)	**CMS Milestone** (calendar year)
1999-2003	R&D Phase	
2000	select Regional Centers	
2002	"Turn on" a prototype functional Regional Center	
2004-2006	Implementation Phase	
2005	Fully operational centers	
2006	Begin operation	
2007	Operations phase	

6.2 User Facilities Work Breakdown Structure

The numbers at the beginning of each of the following items are the Work Breakdown Structure levels. Level 1.0 is the User Facilities Subproject.

1.1 Tier 1 Regional Center at FNAL

This item covers the capital equipment and ongoing operational requirements of the Tier 1 Regional Center. Hardware will be acquired over a three year period from 2004 to 2006, in order to spread the cost over several fiscal years, to provide some reduced capacity capability before 2006 for mock data challenges and test beam analysis, and to provide an adequate time period for full system integration. Consistent with these aims, the hardware will be acquired as late as possible to maximize the performance/price ratio of the systems purchased. There will also be ongoing expenses for the required hardware support to ensure full availability of the services and systems provided. Software systems do not share the falling prices expected of hardware, and in some cases have longer lead times for development, and so software system expenses here will start in earlier years.

1.1.1 Hardware

1.1.1.1 Development and Test Systems

These systems support software development and testing of both system and application software, including testing of user jobs on event samples which are too large for the desktop but do not require full production systems. Small-scale systems of each architecture used in production servers will be required.

1.1.1.2 Simulation Systems

These systems support large-scale production of simulated events, including event generation and full detector simulation. This activity will almost entirely be carried out at Regional Centers and local CMS institutions. This is an extremely CPU intensive application, and will make use of inexpensive commodity computers with limited I/O performance.

1.1.1.3 Reconstruction Systems

These systems support the reconstruction of both data and simulated events. Primary reconstruction of data will take place at CERN in quasi real time, but the regional center will perform primary reconstruction of simulated data, and secondary reconstruction (with improved algorithms or calibrations) of samples of events, on both a scheduled and on-demand basis. Special reconstruction of calibration data samples of interest to US subdetector groups will also be carried out here. This is again a CPU intensive application and will use inexpensive commodity computers.

1.1.1.4 Analysis Systems

These systems provide the computing hardware needed for data analysis, the primary function of the Regional Center. These systems must support large numbers of simultaneous users, provide a single system image no matter what physical hardware the user is logged on to that day, and provide high bandwidth input/output and network connections. Currently UNIX SMP (symmetric multi processor) machines are used for this function. It is expected that the I/O and network requirements will result in somewhat higher cost CPU for this function, but as much use as possible will be made of inexpensive commodity computers.

1.1.1.5 Data Access and Distribution Servers

These systems provide the computing, disk caching and network facilities required to provide access to the event data stored in databases, and to supply this data over the network to both local and remote users.

1.1.1.6 Data Storage Systems (Disk, Tape, Robotic Storage)

These systems provide the hierarchical data storage for event data, calibration data, and non-event data necessary for physics analysis. They will include on-line disk, secondary robotic tape storage, secondary human mounted tape storage, and import/export facilities to move data into and out of the various levels of the data storage hierarchy.

1.1.1.7 Database Servers

These systems provide the computing, disk caching and network facilities required to provide access to the non-event data stored in databases, including geometry, calibration, trigger, run condition, and other data bases.

1.1.1.8 Serial Media and Media Copy and Distribution

This system provides the hardware required to copy and distribute data on physical media (not over the network).

1.1.1.9 Information Servers

These systems provide the computing, disk and network resources required to serve information to the collaboration, primarily using the World Wide Web. This information includes documentation, publications, talks, engineering drawings, and other collaboration notices.

1.1.1.10 Desktop Systems

These systems provide desktop computing resources to local users and visitors at the Regional Center.

1.1.1.11 Monitoring and Problem Response

These systems provide the hardware tools to monitor and optimize performance of regional center hardware, and to detect bottlenecks or other problems at an early stage to allow rapid response.

1.1.2 Software Systems

These systems provide the critical software infrastructure for performing CMS computing tasks, other than the infrastructure for physics code provided by the Core Application Software Subproject. In most cases there is some overlap with the CAS Subproject. In general this subproject concentrates on those services related to operations and support of production systems, while the Core Applications Software (CAS) Subproject concentrates on software design and development, initial or non-production software deployment and end user interface aspects. Software development within the User Facilities Subproject addresses those systems needed to operate and generally deploy CMS software. These include enhancements to commercial packages or tools that increase the efficiency and effectiveness of the Regional Center. User Facilities expenses include the commercial software licensing and maintenance fees, the costs associated with developing and supporting the needed extensions to available software to meet the

CMS collaboration needs, and the personnel costs associated with running and supporting the systems and their services. In all cases these projects are tightly connected with the overall CMS software efforts.

In some cases the US has taken a leadership role in these software projects, in others we are one of several collaborators, but in all cases it is critical that we ensure the chosen tools allow for efficient and effective participation by US physicists in collaboration activities.

1.1.2.1 Database and Database Management

These systems provide the database and database management software and services to store, access, and manage CMS event, calibration, and other non-event data. The User Facilities subproject will work as part of the overall CMS collaboration in the selection and/or development of a CMS standard database system. Extensions and enhancements to the base tools will be developed or acquired as needed for effective access to and management of the data. The size of these databases could be very significant and may need to provide a local central repository for data from and stored on behalf of other local centers. Special techniques are needed to ensure efficient query access, update and synchronization of the databases of the Regional Center with the central site at CERN and other local centers. Complete bookkeeping will be part of the services of the databases provided by the Regional Center. Experience gained from the development and support of such systems during CDF and DØ Run II data taking will be valuable input to the development of the services needed by CMS users. Further development to provide a distributed integrated database system that functions across wide area networks will be required. These developments will benefit from the 'Gird' developments in the US and Europe.

1.1.2.2 Data Access Software

These systems allow for optimized access by CMS physicists to the object-collections of interest, drawn from the large quantities of data stored at the Regional Center. The data access software must provide for and schedule the transfer of data between central site at CERN and the Regional Center, and between the Tier 1 Center and other local Tier 2 centers. It is expected that both network and “Federal Express” shipping options will be required to provide the necessary guaranteed bandwidth to the users of the data. These facilities will be developed taking into account the experience gained from operation and support of the CDF and DØ Run II data access systems and the provision of Run II data to the offsite institutions. The data access software must handle prioritization and allocation of resources among competing users and physics groups, and combining or otherwise scheduling data requests for maximum performance. The systems must provide event and run catalogs and support creation of streams of data and selections of events, or object-collections drawn from events.

1.1.2.3 Data Distribution Software

This system concentrates on the problem of providing data access to those CMS users without good network connections, or who require larger amounts of data than can be conveniently delivered over the network. One solution would be for requested datasets to be automatically extracted and written to serial media, and then physically shipped. Again, these systems must provide resource allocation and optimization of requests.

1.1.2.4 Simulation Software

These systems provide the infrastructure for performing event generation and detector simulation on a production basis. Most of this software is application specific, but support must be provided for the generic detector simulation toolkits (presently GEANT4) that are used, as well as for the physics event generator software.

1.1.2.5 Data and Physics Analysis Software

This system provides tools and environments for analyzing and visualizing the event and non-event data. A major component of this system will be the tools selected as overall CMS standards, LHC++ and other such packages. A sufficient set of these systems must be easily and effectively usable and accessible to the Regional Center users and US physicists. These tools must be available on the typical inexpensive commodity desktop. These systems may require extension to support analysis of the data using multiple complementary tools.

1.2 System and User Support

These services and software provide the fundamental infrastructure for the support and use of the Regional Center Facilities. Because we are supporting physicists doing software development, detector design and test beam work during the R&D phase, system and user support services are necessary during all phases of the project. In addition to supporting physicists during the R&D phase, they provide for the efficient and effective use of the Regional Center facilities by the US CMS physicists, other local Tier 2 regional centers, and by CMS collaborators located throughout the world. In each case, the expenses are a mixture of commercial software acquisition and licensing costs, and personnel costs for software development and ongoing support and maintenance.

1.2.1 Document Generation and Access Tools.

Software to support preparation, access to, management of, and distribution of documentation, both of software systems and of CMS hardware and detector systems. Automatic generation of documentation from the code repository is already being used for Run II. Local system and software documentation and distribution is included in here.

1.2.2 Collaborative Tools

The key goal of facilitating physics research by CMS collaborators independent of their geographic location requires development, widespread deployment and systematic support of new tools to facilitate collaboration at a distance. This includes relatively conventional areas such as distributed code management systems and ordinary videoconferencing, to more advanced approaches such as extended virtual and tele-presence products and innovative distant sharing of physics analysis. Work is ongoing in these areas and further R&D is needed.

The UF subproject focuses on the deployment and general support in the US for these tools. Initial and ongoing R&D is being undertaken at Caltech in collaboration with CERN on remote collaboration systems as part of the CAS subproject.

1.2.3 Code Management and Code Distribution Tools

Tools to provide code versioning, code building, code release and code distribution. These tools must support worldwide physicist-developers, and must allow stable reliable software versions coexisting with rapid development of new code. Here again there will be an interface with the CAS subproject, which will cover the initial design, development and packaging of software releases, while the UF deals with their widespread deployment and support for their use in production-mode.

1.2.4 Software Development Environment

Development and installation of the software, which allows US CMS collaborators to develop software on the Fermilab User Facility employing standard software components and practices established for use throughout CMS.

1.2.4.1 Code Installation

This covers the actual CMS code installations from CERN CVS repositories on user facility machines at Fermilab.

1.2.4.2 Code Verification

This covers software development of code used to verify that the Fermilab and Tier2 center installations are correct and are consistent with the CERN installation.

1.2.5 Software Development Tools

Software systems used by physicists in designing, developing and testing their own software, whether it is to be used privately or included in official CMS software packages.

1.2.6 System Administration, Monitoring and Problem Resolution Software

Tools to allow system administration (account creation and deletion, disk quota and allocation, system software installation and update, etc) for the very large number of individual systems required at the Regional Center, and to allow monitoring and problem detection of the systems and automated and efficient response.

1.2.7 User Help Desk and Information Systems

Personnel and tools to allow for efficient consulting and help services at the Regional Center, providing for needs of both local and remote users.

1.2.8 Training Office

1.2.8.1 Coding classes and instruction

One of the roles of the Regional Center will also be to provide training for US CMS collaborators. We have already initiated training courses at Fermilab in the use of the C++ programming language, based on the successful courses already widely used by CDF and DØ, as well as Object Oriented Analysis and Design. In addition to training in C++, Java, software design and documentation and (later) the software development process for large projects, we intend to provide tutorials on the use of CMS-specific software and computing systems.

Training will take the form of packages for self education as well as formal training classes, and will provide for both local and remote users. Remote collaborative systems being developed will include support for multi-site interactive training classes, using packet videoconferencing and associated tools.

1.2.8.2 Software and training workshops

In addition to more formal classes in coding and design, a series of CMS software workshops is envisioned to instruct US CMS physicists in running and developing software at supported systems at Fermilab.

1.2.9 Support for Tier 2 Centers

This item covers support in the design, implementation, installation and operations phase for Tier 2 Centers. Once these centers have been selected and scoped, the WBS for this item will be fully developed.

1.3 Maintenance and Operation

1.3.1 Hardware Maintenance

This item covers licensing and maintenance of the Regional Center and R&D hardware.

1.3.2 Software Maintenance

This item covers installation, licensing and maintenance of the Regional Center supported software.

1.3.3 Systems Operation

This item covers the Regional Center operation and includes media costs, operations staff, and other such expenses.

1.3.4 Infrastructure Support

This item covers management staff, building and site expenses, power and cooling costs, etc.

1.4 Support of Tier 2 Regional Centers

This item covers support provided by the User Facilities Subproject, and in particular by the Tier 1 Regional Center, for US CMS Tier 2 centers. This includes coordinating with Tier 2 Centers in sharing computing tasks, keeping regional copies of datasets, providing software and keeping it up to date, supplying various consulting and training services, and providing communication and coordination with CMS and CERN.

1.4.1 Support of US CMS University Tier 2 Centers

Several US CMS universities have proposed building modest computing centers at their home institutions for regional support of US CMS physicists. Both hardware and software support for those centers are covered in this item.

1.5 Networking

1.5.1 Onsite Networking

This item covers the network design and procurement, installation and commissioning of all necessary networking hardware within the Feynman Computing Center at Fermilab as part of the Analysis Systems and the Data Management and Access Systems, and other networking specifically for US CMS physics analysis. Additional networking may be required as part of Storage Area Networks providing data for analysis. Security measures to guarantee authorized access only will be implemented.

1.5.1.1 Networking Technology Selection and Prototyping

The networking technology will be evaluated and prototype installations will be performed together with R&D work and test installations. Performance and cost results will be assessed. The hardware choices will be reviewed on an ongoing basis to optimize performance and minimize maintenance and replacement cost.

1.5.1.2 Networking Equipment Procurement and Installation

This item covers installations of Local Area Networks (LANs), and the updating and replacement of obsolete equipment LAN equipment.

1.5.1.3 Operation of Local Area Network

This item covers the operation and ongoing maintenance and tuning of the Local Area Network.

1.5.2 Domestic Networking

This item covers US CMS needs for interfacing to the Wide Area Network for remote access to the data storage of the regional center from all US CMS groups doing analysis on CMS data. The resources required are a sizeable fraction of Fermilab's total needs for domestic Wide Area Network connectivity. Security measures to guarantee authorized access only will be implemented.

1.5.2.1 Connection of Wide Area Network to Local Area Network

This item covers possible US CMS contributions to the procurement, installation or commissioning of the necessary hardware to connect the local area networks to the domestic wide area networks, as required for collaborating groups doing physics analysis.

1.5.2.2 Domestic Wide Area Network Connection

This item covers the operation and connection cost.

1.5.3 International Networking: US to CERN (dedicated CMS part)

The network requirements in the User Facilities Subproject are presently confined to the CMS-specific network facilities at the Fermilab site, and interfacing equipment of sufficient speed and functionality to connect the Regional Center to ESNet and other US national networks serving US CMS collaborators. US-CERN networking is assumed to be provided through common facilities provided for the LHC and other major DOE/NSF high energy physics programs, as is currently the case, in order to exploit economies of scale.

This item may be adapted as needed to ensure that US CMS collaborators at their home institutions, as well as at CERN, have adequate connectivity with the Regional Center.

1.5.4 Networking R&D

There are many candidate networking solutions for onsite intra-system communication, and this item covers the researching, prototyping and testing of different commercial networking systems to see which are most suitable for LAN/SAN operation. It also covers participation in the R&D required to ensure high throughput, transparent access to data distributed among CERN and the regional centers.

1.6 R&D Phase Activities

This item covers both hardware and software activities of the User Facilities Subproject up to and including commissioning of the initial prototype regional center in 2002. These include activities carried out primarily at the site of the Tier 1 Regional Center (FNAL), as well as R&D efforts at other institutes important to the design and operation of the Regional Center itself and the overall distributed data access and analysis architecture. There is some overlap of these activities with the Core Applications Software subproject, especially in the software development and testing area. In general this subproject concentrates on those aspects related to deployment, operations and support, while the CAS subproject concentrates on software design and development, preparation of software releases, and end user interface aspects.

1.6.1 Hardware

1.6.1.1 Distributed Data Management and Distribution Test Beds

A key concept in the CMS computing plan is the use of interconnected centers located throughout the world to perform CMS reconstruction and analysis. The MONARC (Models Of Network Analysis at Regional Centers) R&D project [10], a common project endorsed and monitored by the LCB (LHC Computing Board) is actively engaged in studying this issue. US CMS has assumed an important leadership role in MONARC and a number of US CMS scientists are collaborating on this project.

Hardware will be needed to perform tests of MONARC concepts and to measure performance. These tests will provide important inputs to the actual design of regional centers and their interconnections, and to the proposed strategies to be used for distributed data analysis. This WBS item provides the funding for MONARC associated test-bed systems that will be used to measure fundamental parameters of distributed data analysis using various databases (as input for modeling activities described below), and to create small scale distributed data analysis test-bed systems as described in the MONARC project execution plan. The test-bed systems at FNAL will interact with other systems in the US and in Europe over wide area networks, with an emphasis on the next generation networks that are becoming available in the US.

1.6.1.2 Prototype Fully Functional Regional Center by 2002

It is expected that the hardware for the initial Regional Center operation (which will start at initial data taking in 2005) will be acquired over a three-year period, as described above. However, it is necessary to assemble a lower capacity but fully functional regional center at an earlier date to provide proof of concept at an early enough date to allow design modifications in potential problem areas. CMS has a milestone for such prototyping of Regional Center operation in 2002. This WBS item covers the hardware required for these initial tests.

1.6.1.3 Systems to Support Construction Phase Activities

Much of the computing power needed during the construction phase can be supplied by leveraging the use of shared FNAL computing resources. There will be a need for some modest dedicated CMS systems, including support for test beam data acquisition and analysis, certain compute intensive simulation projects, and software development and distribution servers. Some of this equipment will logically be located at other sites (Tier 2 centers and local CMS institutions) to maximize the efficiency of utilization, and to take advantage of existing facilities that may be available at US CMS universities.

1.6.1.4 Import/Export Facility

In order to support the construction phase of the experiment, and in order to research ideas in data handling and management during the R&D phase, the ability to import and export simulated CMS event data is needed as well as systems for archiving terabytes of data. This WBS item covers these three functions and will serve as both a production system for CMS physicists during the construction phase and as an R&D project for data retrieval and archiving for the Regional Center.

1.6.2 Software

These items cover software activities during the R&D phase prior to the commissioning of fully functional prototype Regional Centers in 2002. Costs are primarily for personnel, located both at FNAL and at other CMS institutes, with some costs for commercial software licensing.

1.6.2.1 Distributed Data Management, Access and Monitoring Software

One of the major responsibilities of the Regional Center is the efficient delivery of data of various types to collaboration physicists. Software must be available to access and deliver the data, taking account of experiment assigned priorities and optimizing use of resources; to manage the data stores, allocating space in the various levels of the storage hierarchy; and to monitor the operation and performance of these systems. This software will most likely be based on a commercial object database management system (ODBMS) such as Objectivity/DB, but additional layers of software to provide user access, monitoring and robust operation across wide area networks that are not provided by the commercial packages will also be needed.

This item covers the deployment and use in production-prototypes of the ODBMS and additional software layers. This will leverage work for FNAL Run II that deals with similar problems on a scale that will approach that needed for CMS. The CAS subproject covers the complementary aspects of the early R&D work on the design and conceptual development of the use of the ODBMS for CMS applications, as well as the initial design and development of any software "superstructure" needed for robust operation over networks.

1.6.2.2 Distributed Computing Model Prototyping Software

As described above, the MONARC project is studying the parameters of distributed data analysis architectures with full participation by US CMS physicists. One important component of this study is to model and simulate the CMS analysis process, including a full complement of (modeled) regional and local centers. This item covers the installation, support and use of prototype software, including a selected set of analysis and database applications, in production-prototype systems to test the MONARC concepts and measure key parameters that govern the performance of the distributed systems.

This complements, and must be coordinated with, the efforts to understand and characterize the physics analysis process, and to develop, run and analyze simulations of these systems, which are described in Section 8.2, under CAS Work Breakdown Structure item 2.3.

1.6.2.3 Production System infrastructure software

This software covers fully integrating the CMS software into a production system. Examples are batch systems, archiving system, automated job submissions, etc.

1.6.2.4 Data import/export software

In order to support US CMS during the construction phase, and as R&D for a Tier 1 Regional Center, the problem of media exchange and non-networked data import/export must be attacked. This WBS covers development of software for importing/exporting data from/to US CMS institutions and from/to CERN.

1.6.2.5 Data archival/retrieval software

This covers software necessary to archive and retrieve data and databases in a mass storage system.

6.3 User Facilities Budget and Personnel Requirements

The scope of the Regional Center is taken from the results of MONARC and is consistent with CMS estimates. It represents 20% of CERN's projected capability for a single experiment. This coincides nicely with the US fraction of CMS and hence represents a fair contribution. The hardware costs are based on the CMS Computing Technical proposal[1], Fermilab Run II experience, CMS estimates, and present Fermilab expenditures. The CPU requirements include an estimate of the overhead engendered by the use of an object database to store data (network access and database processing).

It is difficult to set the personnel needs of the UF Subproject since there is as yet no CMS Technical Design Report for Computing. As a first estimate, the profile and FTE requirements have been based on experience with the Fermilab Run I and Run II computing and software projects, scaled appropriately. We will continue to refine these estimates as we increase our understanding of the needs of CMS, but we do not expect the final results to be significantly different than those shown.

The cost and personnel for the fiscal years 1999-2006, and for operations in the years 2006 and onward, are listed below. To summarize:

- During the R&D phase of the project (roughly 1999-2003), we will purchase test bed and prototype systems for studying Regional Center architectures while also supporting ongoing US CMS simulation, software development and test beam analysis activities. The total hardware cost for R&D activities is $2.22M. During this period, the number of Fermilab CD personnel working on user facilities activities will increase from 3 (1999) to 13 (2003). The total anticipated new personnel cost for the R&D phase is $3.4M.
- The total hardware cost for the Regional Center itself is $9.2M. It will be purchased during 2004-6. The three-year purchase period is the same length as that for the Run II systems. We have estimated that an additional $100k per year will need to be spent on software licenses, and $1.8M for networking, during the implementation phase. During this time the number of Fermilab CD personnel working on user facilities activities will rise from 13 (2003) to 35 (2006). The latter number is estimated from the Run II experience of the staff required to

support a similar number of physicists in CDF or DØ. The total cost of new personnel for the Regional Center itself during 2004-2006 is expected to be $12.4M.

- During the operations phase, $3.1M will need to be spent on hardware acquisitions per year (to replace and augment the facility). Networking will cost $1.2M per year and licenses $120k. Continuing costs for the staff supported by User Facilities subproject funds will be $5.0M/year.

We note that our CPU cost estimates are higher than some others are; this reflects the choice of relatively expensive SMP machines in the Run II plan from which the costing is derived. If the computing model permits the use of mostly commodity CPU (as was not the case for Run II), the CPU cost might be reduced. It should be emphasized however that FNAL (as well as SLAC) experience indicates that a configuration that includes a core of facilities with higher I/O performance is more cost-effective than an architecture based wholly on loosely coupled computing "farms", once the manpower requirements and costs are properly taken into account.

The detailed resource estimates are given in three tables, which follow.

Table 1 shows the FTE profile for the Level 2 WBS categories in the User Facilities Subproject and their distributions at Level 3. In the near term, we anticipate a need for 7 FTEs in 2000, 9.5 in 2001, 11.5 in 2002, and 13 in 2003. We intend to leverage 3, 5,6 and 6 from existing Fermilab staff, and hence expect four new positions will be needed in 2000, 0.5 additional position in 2001, one additional position in 2002, etc. The numbers include a contingency factor of 25% for the years 2004 onwards to handle difficulties and guarantee the completion of the User Facilities Subproject by the start of data taking in 2006 and a successful startup. For 1999-2000 no contingency is taken into account, while for 2001-2003 a contingency factor of 10% is included

Table 1 also shows the anticipated costs of new personnel. The average market base salary for suitably skilled Computing Professionals in the Chicago area at Fermilab, at the end of 1999, is about $78k/year. Given 28.5% for benefits, 30% for overhead, and adding travel costs, infrastructure and software licensing costs, the fully encumbered average salary for a Computing Professional is approximately $154k/year. The cost in the table is corrected for inflation starting in 2001.

		1999	2000	2001	2002	2003	2004	2005	2006	2007
1.1	**Tier 1 RC**	**0**	**0**	**0**	**0**	**0**	**8**	**18**	**20**	**17**
1.1.1	Hardware	0	0	0	0	0	4	9	9	7
1.1.2	Software Systems	0	0	0	0	0	4	9	11	10
1.2	**System and User Support**	**1**	**1.25**	**1.5**	**1.5**	**2**	**3.5**	**5**	**6**	**8**
1.3	**Maintenance and Operation**	**0.5**	**1**	**1**	**1.5**	**1.5**	**3**	**5**	**5**	**6**
1.4	**Support for Tier 2 Centers**	**0**	**0.25**	**0.5**	**1**	**1.5**	**1.5**	**2**	**2**	**2**
1.5	**Networking**	**0**	**1**	**0.5**	**0.5**	**1**	**1**	**2**	**2**	**2**
1.6	**R&D Phase Activities**	**1.5**	**3.5**	**6**	**7**	**7**	**6**	**3**	**0**	**0**
1.6.1	Hardware	0.5	2	3	3	3	3	2	0	0
1.6.2	Software	1	1.5	3	4	4	3	1	0	0
User Facility (total FTE)		**3**	**7**	**9.5**	**11.5**	**13**	**23**	**35**	**35**	**35**
of which										
	FNAL Base Program	3	3	5	6	6	7	8	8	8
	CMS Computing Project Request	0	4	4.5	5.5	7	16	27	27	27
User Facitlity (total personnel cost in $M)		**0.46**	**1.08**	**1.51**	**1.88**	**2.19**	**3.87**	**6.07**	**6.25**	**6.44**
of which										
	FNAL Base Program	0.46	0.46	0.79	0.98	1.01	1.18	1.39	1.43	1.47
	CMS Computing Project Request	0.00	0.62	0.71	0.90	1.18	2.69	4.68	4.82	4.96

Table 1: Personnel Requirements for the User Facilities Subproject, summarized for the higher level WBS items. The numbers include a contingency factor of 25% from years 2004 onwards to guarantee the completion by start of data taking in 2006 and a successful startup period. For the years 1999-2000 no contingency is taken into account, for the other years a factor of 10% is included. The cost is corrected for inflation starting in 2001; the official DOE escalation index is used.

Table 2 shows the evolution of installed capacity for CPU, disks, tape, and robotic mass storage for the implementation and operation phases of the Tier 1 Regional Center.

FY	2000	2001	2002	2003	2004	2005	2006	2007 (CMS operations)
CPU (SI95)	-	-	-	-	7k	30k	70k	110k
Disk (TB)	-	-	-	-	10	40	100	150
Tape (PB)	-	-	-	-	0.06	0.25	0.6	0.9
Robots	-	-	-	-	1	1	1	1.33

Table 2: Total Installed capability by year, for Implementation and CMS Operations Phases of the Tier 1 Regional Center. In order that spending is roughly equal in each year, we plan to acquire 10% of the disk and CPU in 2004, 30% in 2005, and 60% in 2006. We will buy one tape robot in 2004,

and then budget to buy an additional one every three years during operations, hence the 0.33 robots apparently purchased in 2007.   Test-bed and prototype systems for 1999-2003 are not included.

Finally, Table 3 shows a summary by year of all hardware, R&D, networking, and software licensing costs.

FY	1999	2000	2001	2002	2003	2004	2005	2006	Total	2007 (CMS operations)
CPU ($M)	-	-	-	-	-	0.91	2.46	3.08	6.45	2.13
Disk ($M)	-	-	-	-	-	0.20	0.40	0.54	1.14	0.44
Tape and robot ($M)	-	-	-	-	-	0.90	0.08	0.02	1.00	0.56
Total Regional Center						2.01	2.94	3.64	8.59	3.13
R&D systems:										
at Fermilab	0.00	0.28	0.51	0.75	0.75				2.29	
elsewhere	0.00	0.10	0.07	0.05	0.05	0.06			0.33	
Networking within the US							0.52	0.57	0.73	1.82 1.24
Licenses							0.11	0.11	0.12	0.34 0.12
User Facility cost (except personnel)	0.00	0.38	0.58	0.80	0.80	2.7	3.62	4.49	13.37	4.49

Table 3: User Facilities Subproject costs for the R&D, implementation, and operations Phases. The hardware costs for R&D systems cover testbed and prototype systems for the Tier 1 regional center, support of ongoing US CMS simulation, software development and test beam analysis activities at Fermilab, and the provision of a disk pool (1TB/year) for ODBMS R&D and a DLT tape system for Monte Carlo event distribution at Caltech. The cost is corrected for inflation starting in 2001; the official DOE escalation index is used.

7. High Level Milestones, Work Breakdown Structure and Budget for the Core Applications Software Subproject

7.1 Core Applications Software High Level Milestones

7.1.1 Schedule Considerations

The schedule is defined taking into account a number of constraints: the ongoing needs of the physicists involved in the detector and trigger design optimization and construction; the research and development required, particularly for large network-distributed storage and computing systems; and the availability of resources, especially manpower. The schedule reflects an iterative development strategy, which aims to provide continuously working software while simultaneously evolving into software systems with the functionality and performance that is ultimately required.

7.1.2 Software Milestones

The first key element of the plan is the management of the transition from a procedural software paradigm, based on Fortran and C, to an Object Oriented paradigm based on C++. This is necessary in order to manage the complexity of the task and to maintain consistency with prevalent software practices elsewhere in the research and commercial sectors. The transition will need to be carefully managed as detector design, simulation, and test beam activities continue to require continuous computing support. The software following the transition will provide a powerful and user-friendly environment for physicists to carry out physics studies related to the optimization of the detector and trigger. It should be emphasized that the additional effort initially required to make this transition will be rapidly compensated by the benefits of using more modular and manageable OO software.

Table 4 shows the major milestones for the CMS Software sub-project, which have been approved by the CERN LCB committee. They reflect the iterative software development process, which has four phases. The first phase, which was completed at the end of 1998, is the proof of concept for the key elements of the software architecture proposed in the CMS Computing Technical Proposal [1]. The second phase, which is currently in progress, covers the development of functional prototypes of the various software modules, which are now being exercised using large samples of simulated events and test beam data. This software has been, and is being, used for detailed studies of the detector and its sensitivity to various physics processes. The third phase involves the development of the prototypes into fully functional software while the fourth phase is the preparation of the software for production and the exercising of the complete system, in conjunction with the pre-production hardware systems. The milestones associated to the database reflect its central importance in the CMS Computing Model.

CORE SOFTWARE	1997	1998	1999	2000	2001	2002	2003	2004	2005
End of Fortran Development		●							
GEANT4 Simulation of CMS		❶	❷		❸		❹		
Reconstruction/Analysis framework		❶ ❶	❷ ❷		❸	❸	❹	❹ ❹	
Detector reconstruction									
Physics Object Reconstruction			❶	❷		❸			
User Analysis Environment		❶		❷		❸		❹	

DATABASE	1997	1998	1999	2000	2001	2002	2003	2004	2005
Use of ODBMS for test-beam data	●	❶	❷		❸			❹	
Event storage/retrieval in ODBMS									
Data organization/access strategy		●							
Filling at 100MB/s			●						
Simulation of data access pattern				●					
Integration of ODBMS and MSS				●					
Choice of vendor for ODBMS					●				
Installation of ODBMS and MSS						●			

Table 4: Milestones for CMS Computing and Software Project (1- proof of concept; 2-functional prototype; 3-fully functional prototype; 4- production software)

7.2 Core Application Software Work Breakdown Structure

Since this subproject is intimately connected to the comprehensive CMS software project, we include the Work Breakdown Structure for that project in Appendix A. Here, as we present the WBS for the Core Applications Software Project, we provide pointers to the CMS Software WBS where appropriate.

2.1 ORCA – Detector Reconstruction (CMS WBS 1.3.10)

This item covers software engineering and professional programming support for the US activities connected to the ORCA (Object Reconstruction for CMS Analysis) project. As a principle CMS tries to ensure that software responsibilities are assumed or at least shared by the groups responsible for the sub-detector construction. This is reflected in the WBS for this task, described below.

2.1.1 Calorimetry

This item covers software engineering and professional programming support for the US activities connected to ECAL and HCAL. The ECAL and HCAL software are both part of a unified “Calorimetry” subsystem with much of the software and interfaces being common for the two systems.

2.1.2 Endcap Muons

This item covers software engineering and professional programming support for the US activities connected to the endcap muon systems. It is part of the overall “Muon” software sub-system, which encompasses both the barrel and endcap systems. Specific responsibilities include the muon track fitting in this detector and the implementation of the mechanisms for creating and processing digitizations.

2.1.3 Tracker

This item covers software engineering and professional programming support for the US activities connected to endcap pixel software, detector description and simulation studies. Where the very different geometries permit, software is shared or reused for both barrel and endcap pixels.

2.1.4 Trigger

This item covers software engineering and professional programming support for the US activities connected to the simulation of the level-1 trigger, which is vital for an understanding of the physics capabilities of the detector. This includes responsibility for simulating the regional calorimeter trigger in ORCA, for implementing the simulation of the calorimeter trigger signals required by the level-1 trigger simulation, and for the simulation of the Forward Muon CSC (Cathode Strip Chamber) trigger.

2.1.5 Detector and Event Visualization

This item covers software engineering and professional programming support for the US activities connected to the provision of interactive 3D detector and event visualization software. Much of this software is generic to CMS, and indeed HEP, and is therefore part of the “User Analysis Environment” task described below. In addition, a significant amount of ORCA-specific software is required to produce graphical objects from reconstructed objects and to support their interactive manipulation.

2.1.6 Examples

This item covers software engineering and professional programming support for the US activities connected to the provision of a set of documented and working examples of simple analyses based on ORCA software. This work is related to the “User Analysis Environment” task described below.

2.1.7 Persistency

This item covers software engineering and professional programming support for the US activities connected to support within ORCA of persistent storage in an ODBMS for objects of subsequent interest to physicists. These include hits and digitizations, reconstructed tracks and clusters, etc. This provides crucial input to the “Physics Reconstruction and Selection and Higher Level Triggers” groups. By avoiding the need to repeat the time-consuming ORCA reconstruction step the process of optimizing higher-level algorithms will be speeded up significantly. The general mechanisms of persistency depend on the ODBMS and CARF subtasks. This work combines the software aspects of implementing persistency with studies of the computing and distributed system aspects (described below), thereby ensuring that the solutions will work well in practice.

2.2 User Analysis Environment (CMS WBS 1.3.12)

This item covers software infrastructure and tools for performing user analysis of real and simulated CMS data. It includes the following tasks: interactive data analysis and presentation including histogramming; interactive detector and event visualization; (graphical) user interfaces, and statistical and numerical analysis.

2.2.1 Interactive Visualization

This item covers software engineering and professional programming support for the US activities connected to developing the ability to interactively visualize the CMS detector elements and their response to particle interactions. This is of paramount importance in the design, construction, and data analysis phases of the experiment. Detector and event visualization programs are used for the design and debugging of the software, the detector, and the readout, in the evaluation of event reconstruction algorithms, and in producing understandable images for communicating complex concepts to the physics community and the general public. This software will be deployed for simulated, real, and test-beam events in the context of ORCA and physics analysis programs.

2.2.2 Physics Analysis Tools

This item covers software engineering and professional programming support for the US activities connected to developing "Physics Analysis Tools". These include software for interactive data analysis and presentation, statistical and numerical analysis, and histogramming. It must initially provide at least the equivalent functionality of PAW but be newly implemented using maintainable and extensible OO software. Then it must subsequently evolve to fully exploit the much higher flexibility of the OO data model compared to the traditional solutions.

This task is related to the LHC++ project which is replacing the procedural CERNLIB FORTRAN/C libraries with a suite of OO software (predominantly C++) with roughly equivalent functionality. The scope of the LHC++ project covers the following: foundation level class libraries; mathematical libraries; graphical libraries; visualization tool-kits, data analysis, and, and histograms; event generators; detector simulation (GEANT4); and object persistency (RD45). The project exploits solutions based on commercial software where appropriate and affordable.

2.3 ODBMS -- Object Database Management System (CMS WBS 1.3.15)

2.3.1 Persistent Object Management

This item covers the development, testing, commissioning and operation of the Object Database Management system (ODBMS) to be used by CMS and the other LHC experiments to manage persistent objects, as well as some of the associated architectural issues in the experiments' object oriented software. The database management system, which will be chosen by the end of 2001, will provide the mechanisms for data persistence, internal relationships, wide-area heterogeneous' distribution, and concurrent read-write control. It is intended that the ODBMS, with enhancements for the robust operation across wide area networks, also provide transparent access (independent of data location and the storage medium as seen from the user's code) to persistent objects stored in a network-distributed database federation.

The "Objectivity/ DB" product has been chosen as the current focus of work by CERN, CMS and ATLAS, as well as BaBar and other large experiments. Much of the work on object persistency has been carried out in the context of the RD45 joint project[6] at CERN and in the Caltech/CERN/HP/FNAL GIOD project[7]. RD45 and Caltech (HEP and the Center for Advanced Computing Research) have also surveyed, and investigated other competing database systems, beginning sufficiently in advance of the final ODBMS choice to provide useful input. The storage in robotic tape systems (tertiary storage) of infrequently accessed data will be accomplished using the High Performance Storage System (HPSS)[8] .

2.3.2 Networked Object Databases

This item covers software engineering and professional programming support for the US activities which address specific technical issues associated to the operation of Petabyte-scale object databases over local- and wide-area networks. Currently, this is accomplished by participation in GIOD and the DOE/NGI and HEP-funded Particle Physics Data Grid[9], which are complementary to and compatible with the aims of the RD45 and MONARC[10] projects.

2.3.3 MONARC

This item covers software engineering and professional programming support for the US participation in the MONARC Project (Models of Networked Analysis at Regional Centres). This project studies network-distributed computing architectures, data access and analysis strategies, and data management systems that are the major components of the LHC Computing Models and the ways in which the components interact across networks. The approach taken by MONARC is to develop discrete-event and other simulations of the computing facilities at each site, the networks and their handling of the traffic flow, the users' demands on the network, and high-level representations of key components such as the database system and hierarchical storage manager. While the software-development aspects of this work are part of the Core Applications Software subproject, other activities associated with MONARC are being carried out within the User Facilities subproject [3].

2.4 Software Support

This activity supports physicists in the US who wish to develop and use CMS software in their home institutes. This task includes the development, deployment and subsequent support of the CMS software sub-systems, associated tools, and the developers' and users' environment at US institutes, together with associated training in their use. This task also covers software engineering and professional programming support to US physicists participating in the detailed evaluation of sub-detector and trigger systems and the overall CMS physics capabilities. These activities include the assessment of the input to the higher level triggers coming from the Level 1 hardware trigger as well as the evaluation of the Level 2 and Level 3 selection algorithms. Special requirements on the software arise from CPU and bandwidth restrictions in the Level 2 /Level 3 trigger.

Whereas the Core Applications Software Subproject concentrates on development and initial deployment and support, the User Facilities Subproject is responsible for general support of stable production software. For example, the Core Applications Software Subproject covers the design, development, and test deployment of systems for collaboration over networks, while the User Facilities Subproject includes the general deployment and support for such systems once they are reasonably stable. Similar categorization of responsibilities will occur whenever new types of software systems are developed (within the Core Applications Software Subproject) and then brought into production use (in the User Facilities Subproject).

2.4.1 Software Support Tools (CMS WBS 1.3.3)

This sub-task covers the US CMS contributions to the development of generic tools for use not only by US CMS but also throughout the wider CMS community.

2.4.1.1 System Configuration Management

This item covers software engineering and professional programming support to help develop the process and supporting programs for software configuration management, build procedures, and release mechanisms in the context of various compilers, platforms, operating systems, and applications. This item also covers support of the maintenance load for various parts of the CMS software on specific operating systems based on a division of labor among all CMS collaborating institutions.

2.4.1.2 Collaborative Working Tools

This item will provide software engineering and professional programming support for the development of network-based systems for videoconferencing and other tools for computer-supported collaborative work, which are required to support daily communications within CMS, collaborative development of the software, as well as the execution of the data analysis. These tools must be flexible enough to take advantage of new networks as the situation evolves.

2.4.2 Tools to Address Specific Issues for US CMS (no overall CMS WBS)

Despite the widespread use of standards and common solutions, it is expected that there will be some limited areas where the US CMS software community requires some additional software support than that provided by the collaboration at large. Examples of areas needing such support could be support of particular operating systems or hardware platforms used widely in the US but not elsewhere in the collaboration, specific requirements imposed by computer security considerations, or other software requirements due to particular hardware and network configurations present in the US. Moreover, some general-purpose tools could give better performance when optimized for dedicated US use. This subtask will provide the software and engineering and development resources to meet these needs.

2.4.3 Professional Software Engineering Support (no overall CMS WBS)

This subtask provides professional software expertise to support US CMS physicists developing simulation, reconstruction and physics analysis software (which is not already covered by the ORCA, IGUANA, and ODBMS activities supported by WBS elements 2.1, 2.2, and 2.3). Even though this software is not a formal part of the US CMS Software and Computing Project, it will be highly cost-effective to have a single pool of software experts who can, by being shared among different tasks, communicate CMS software standards throughout the US CMS community. Tasks to be performed include mentoring, code reviews, and provision of examples and templates, and software tools.

7.3 Core Applications Software Budget and Personnel Requirements

The resources required for the CMS Software subproject and for the US CMS Core Applications Software subproject are estimated using the resource-loaded WBS, described in the Appendix. This WBS will continuously evolve to meet the changing needs, schedule, and availability of resources. The dominant resource required is professional software manpower, which is obtained by filtering the resource-loaded schedule for such personnel (*i.e.* no physicist manpower is included). The CMS Software sub-project requires a total of 21.5 FTE's in 1999 of which 14.5 are currently deployed. The total for CMS will plateau at 33.3 FTE's in 2003 in preparation for the operation of production software, in conjunction with the production hardware systems, in 2005. The incremental requirement is estimated to remain approximately flat for several years after 2005. This is to permit the shakedown and optimization of the operational system with a real and increasing data set and significantly increased use activities. The personnel requirements of the US CMS Core Applications Software sub-project that are directly associated to activities in the CMS Software sub-project are estimated assuming a 25% US contribution to the total. In addition, the US CMS Core Applications Software sub-project includes personnel dedicated to supporting software for US physicists.

The personnel requirements for the US CMS Core Applications Software sub-project are shown in Table 5. The support personnel have been rounded to yield integer FTE's in the total requirement. This procedure is consistent with the US CMS strategy of dedicating approximately 25% of each FTE to an appropriate support task to ensure their contact with the physicist developers and users (rather than assigning dedicated personnel full-time to the support task). Although the overall scope of the US CMS Applications Software subproject is bounded by this canonical scaling law, specific tasks are assigned to US CMS individuals in the WBS. By filtering explicitly on US-tagged personnel, consistency is maintained with the over all US CMS profile and responsibilities are adjusted accordingly as a function of time. The US CMS Core Applications Software subproject cost profile is obtained assuming:

- The annual salary cost, including overhead and fringe benefits, is $154k / FTE. This is the average of the actual 2000 US CMS requests. The base salaries for these are based on median salaries for such personnel based on previous experience and industry surveys, with allowances for geographical location and skills. The rates of overhead and fringe assumed are those currently in effect at the institutions involved.
- Each FTE requires a desktop system costing 6k$, with a lifetime of 3 years.
- Each FTE requires 10k$ / year to cover associated items such as travel, training and documentation, and software licenses (not covered elsewhere by CMS or US CMS).
- The contingency requested is 0% for the years 1999 and 2000, 25% for the years 2004 onwards and 10% elsewhere.
- The cost is corrected for inflation starting in 2001; the official DOE escalation index is used.

The costs for the US CMS Core Applications Software subproject are shown in Table 5. The cost requirement rises from $1.00 M in 1999 to a plateau of $2.8M in 2005, yielding a total cost for 1999–2005 of $14.1M. The year 2006 is shown to indicate the required resources during each year of operation of the CMS experiment. The numbers for contingency are 25% from years 2003 onwards to guarantee the completion by start of data taking in 2005 and a successful startup period. For the years 1999-2000 no contingency is taken into account, for the other years a factor of 10% is assumed to be adequate.

	FY1999	FY2000	FY2001	FY2002	FY2003	FY2004	FY2005	FY2006	Total	FY2007
CAS FTE request	7	9	10	11	11	12	13	13		13
CAS FTE Contingency	0	0	1	1	1	3	3	3		3
CAS [$M]	**1.00**	**1.38**	**1.58**	**1.79**	**1.79**	**2.01**	**2.25**	**2.32**	**14.12**	**2.38**
CAS Contingency [$M]	**0.00**	**0.00**	**0.16**	**0.16**	**0.16**	**0.51**	**0.52**	**0.54**	**2.05**	**0.56**
CAS Total [$M]	**1.00**	**1.38**	**1.74**	**1.95**	**1.95**	**2.52**	**2.77**	**2.85**	**16.16**	**2.94**

Table 5: Estimated requirements of software personnel (FTE) for the US CMS Core Application Software subproject, including US-specific support and the cost for the sub-project, including personnel cost and associated support, such as travel, desktops, training and software licenses. The numbers for contingency are 25% from years 2004 onwards to guarantee the completion by start of data taking in 2006 and a successful startup period. For the years 1999-2000 no contingency is taken into account, for the other years a factor of 10% is assumed to be adequate. Contingency numbers for FTEs are rounded to integer values.

8. High Level Milestones, Work Breakdown Structure and Budget for Overall Project

8.1 High Level Project Milestones

The high level milestones for the US CMS Software and Computing Project are chosen to be consistent with the overall high level milestones of the CMS Software Plan and the LHC/CMS schedule. These are shown in sections 6.1 and 7.1 for the User Facilities Subprojects and the Core Applications Software, respectively. The overall CMS milestones for software are shown in Table 4.

8.2 High Level Work Breakdown Structure

In addition to the two main elements of the project, the Core Applications Software Subproject and the User Facilities Subproject, there is a subproject which supports the Level 1 Project Manager's activities.

8.2.1 The User Facilities Subproject

This subproject is described in section 2.2. Section 6 gives the WBS, milestones, and budget for this subproject.

8.2.2 The Core Applications Software Project

This subproject is described in section 2.1. Section 7 gives the WBS, milestones, and budget for this subproject.

8.2.3. Project Office

This item provides personnel and operating funds to support the management of the project. It includes a staff of 1/2 FTE administrative assistant, 1/2 FTE Budget officer/planner, and 1 FTE assistant project manager. The budget for travel for the L1PM and for support of reviewers, special consultants, etc is also under this item. The budget is currently absorbed in the User Facilities Subproject budget.

8.3 Budget Summary

Table 6 below lists the budget for the whole US CMS Software and Computing Project. This table summarizes the resources as described in Table 3 and Table 5 and the cost of the project office.

The numbers for contingency for CAS and UF FTE requests are 25% from years 2004 onwards to guarantee the completion by start of data taking in 2006 and a successful startup period. For the years 1999-2000 no contingency is taken into account, for the other years a factor of 10% is assumed to be adequate. For the UF hardware request the contingency is included in the numbers quoted.

	FY1999	FY2000	FY2001	FY2002	FY2003	FY2004	FY2005	FY2006	Total	FY2007
CAS [$M]	1.00	1.38	1.58	1.79	1.79	2.01	2.25	2.32	14.12	2.38
CAS Contingency [$M]	0.00	0.00	0.16	0.16	0.16	0.51	0.52	0.54	2.05	0.56
CAS Total [$M]	**1.00**	**1.38**	**1.74**	**1.95**	**1.95**	**2.52**	**2.77**	**2.85**	**16.16**	**2.94**
UF FTE request [$M]	0.00	0.62	0.56	0.74	1.01	2.02	3.81	3.93	12.68	4.05
UF FTE contingency [$M]	0.00	0.00	0.16	0.16	0.17	0.67	0.87	0.89	2.92	0.92
Project Office	0.00	0.00	0.32	0.33	0.34	0.34	0.35	0.36	2.02	0.37
UF FTE total [$M]	**0.00**	**0.62**	**1.03**	**1.23**	**1.51**	**3.03**	**5.03**	**5.18**	**17.62**	**5.33**
UF Hardware	**0.00**	**0.38**	**0.59**	**0.80**	**0.80**	**2.70**	**3.62**	**4.59**	**13.48**	**4.49**
Total (without Tier2) [$M]	**1.00**	**2.37**	**3.44**	**4.76**	**4.76**	**9.19**	**11.42**	**12.62**	**49.56**	**12.76**
Tier 2 FTE [$M]		0.21	0.56	0.70	0.70	0.80	1.13	1.16	5.26	1.19
Tier 2 Hardware [$M]		0.29	0.44	0.62	0.62	2.06	2.72	3.44	10.19	3.37
Total [$M]	**1.00**	**2.87**	**4.44**	**5.30**	**5.58**	**11.11**	**15.27**	**17.22**	**62.79**	**17.32**

Table 6: Budget Summary of US CMS Software and Computing Project. Amounts are given in units of $M. The costs shown for Tier2 centers in this table are for staff and hardware located at the Tier 2 centers themselves. A more detailed description of Tier 2 centers is given in section 2.2.2 (see also ref. 16).

REFERENCES

6 RD45 - A Persistent Object Manager for HEP. See http://wwwinfo.cern.ch/asd/rd45

7 J. Bunn, Globally Interconnected Object Databases (GIOD) Project Summary Report. Internal CMS Note IN/1999-044, October 1999. See also http://pcbunn.cacr.caltech.edu

8 A. Shoshani et al., Storage Management for High Energy Physics Applications, Computing in High Energy Physics 1998 (CHEP 98). See http://www.lbl.gov/~arie/papers/proc-CHEP98.ps and http://gizmo.lbl.gov/sm/.

9 The Particle Physics Data Grid involves all of the US HEP laboratories, JLAB, Caltech, Wisconsin (Computer Science) and the San Diego Supercomputer Center. See http://www.cacr.caltech.edu/ppdg.

10 M. Aderholz et al., Models of Networked Analysis at Regional Centres for LHC Experiments (MONARC) Project Execution Plan, LCB 99-5, June 1999; see also http://www.cern.ch/MONARC

Appendix: CMS Software Schedule and Work Breakdown Structure

The Core Applications Software is closely aligned with the overall CMS software. In order to understand how the US project relates to all of CMS software development, we present here the current CMS Software WBS at Level 3. The WBS for the Core Applications Software project references the CMS WBS numbers shown here wherever appropriate.

The CMS WBS has the CMS Software Project as a Level 2 Project designated as 1.3. The Level 3 software subtasks are described in detail below.

CMS Schedule and Milestones

Table 4 above shows the major milestones for the CMS Software subproject, which have been approved by the CERN LHCC. They reflect the iterative software development process, which has four phases. The first phase, which was completed at the end of 1998, is the proof of concept for the key elements of the OO software architecture proposed in the CMS Computing Technical Proposal. The second phase, which is currently well underway, covers the development of functional prototypes of the various software modules that are being exercised using large samples of simulated events and test beam data. This software will be used for detailed studies of the detector and its sensitivity to various physics processes. The third phase involves the development of the prototypes into fully functional software while the fourth phase is the preparation of the software for production and the exercising of the complete system, in conjunction with the pre-production hardware systems. The milestones associated to the database reflect its central importance in the CMS Computing Model.

CMS Work Breakdown Structure

WBS 1.3.1 CMSIM

CMSIM is the Fortran-based CMS simulation package. Since new development has essentially ceased, CMSIM will need to be supported for physics studies until GEANT4 and the corresponding OSCAR simulation software have comparable functionality.

WBS 1.3.2 Software Process

This item covers the process associated with the definition, design, development, documentation, integration, verification, deployment, and maintenance of the CMS Software with the aim of improving efficiency and quality[1]. The pragmatic "Cyclic Life Cycle" model which emphasizes continuous improvement following ISO/IEC 15504 (SPICE) [2] has been adopted.

Items which have been implemented include the structure of the ORCA software repository and the strategy for ORCA development, release, and testing [2] and the CMS coding rules[3], guidelines[4], and style-guide[5]. Processes currently under construction include the more comprehensive assessment of user requirements, checking of software dependencies between software subsystems and packages, automated checking of coding rule and style conformance, and more comprehensive integration of test suites and examples. Processes to be implemented more rigorously in future include: more formal and uniform design procedures; documentation of code design, implementation, and usage; and problem reporting, tracking and resolution mechanisms.

WBS 1.3.3 Software support

This item covers support for the CMS software and environment including the associated code management systems, version control mechanisms, and the code release, distribution, and build procedures [1]. The current solution for CMS code is to organize the code into major tasks (ORCA, OSCAR, User Analysis Environment, etc.), which are subsequently divided hierarchically into sub-systems (Tracker, Calorimetry, etc.), each of which contains a number of smaller software "packages" (each of which corresponds to a software library). The CVS system, which is prevalent in HEP and elsewhere, is used to manage the access, change, and version control of the software repositories. The build and distribution system uses SCRAM, a product developed by CMS. An important aspect of this task is the porting, technical verification, and support of CMS and external code on multiple systems. This task also includes the development and support of network-based collaborative working tools such as videoconferencing, which are invaluable to the highly distributed members of any HEP collaboration.

WBS 1.3.4 Information Systems

This item covers the construction, operation, and maintenance of the collaboration information systems. These are centered around the CMS WWW server and include: management of news, agendas, and minutes in a uniform fashion for the complete hierarchy of CMS groups; the CMS collaboration members database; and storage, access, and archival of documents, including internal working documents, technical notes, and collaboration official documents.

WBS 1.3.5 CARF -- CMS Analysis and Reconstruction Framework

The foundation of the CMS software is a professionally engineered framework into which physics modules may be inserted. The "CMS Analysis and Reconstruction Framework" known as CARF [6], aims to provide the user with an intuitive and flexible means of performing analysis and reconstruction tasks. CARF should support standard operations performed by physicists, including: accessing the CMS data store; selection and classification of events; the ability to apply calibration algorithms; invocation of standard or custom-built reconstruction algorithms; interactive visualization and physics analysis systems with presentation-quality output; and creation of user-defined transient or persistent objects. The user of CARF will be able to choose what specific actions to perform and which specific algorithms to use even at run-time or interactively. CARF will use an ODBMS to store objects persistently and will provide classes and methods to shield end-users from details of the ODBMS storage and access mechanisms, mostly in terms of C++ smart references and smart iterators. The underpinnings to ensure that the ODBMS operations of CARF function as planned are being addressed through the ODBMS task described below.

WBS 1.3.6 Event Generators

This item covers the software required to support external physics event generator programs in the CMS environment. In particular it includes software for the management of input parameters in a consistent fashion and procedures for storing output events with a standard format for subsequent use by, for example, OSCAR and the Fast Simulation software. This task does not cover the software of the generator programs themselves.

WBS 1.3.7 Detector Description

This task will provide an environment for creating, manipulating, and using the parameters describing the CMS detector in a consistent manner [7]. In particular, it covers the geometrical description of the detector elements at various levels (full engineering detail, full GEANT detail, fast simulation, trigger tower geometries, etc.), associated material properties, magnetic field map, etc. The Detector Description sub-system will serve a number of clients, including OSCAR, the Fast Simulation, ORCA, the Calibration, and the User Analysis Environment.

WBS 1.3.8 Fast Simulation

This item covers fast simulation of particles traversing the CMS detector and the response of the detector elements, readout electronics, and triggers [8]. Parameterizations will be verified with OSCAR simulations, test-beam analyses, and ultimately real CMS data. The Fast Simulation should use the Common Detector Description and be consistent with OSCAR, ORCA, and the Physics Object Reconstruction software to permit flexible the use of varying degrees of simulation detail according to the requirements of the user.

WBS 1.3.9 OSCAR -- Detector Simulation

This item covers the full simulation of particles traversing the CMS detector using the GEANT4 Toolkit, including the use of an appropriate detector description, tuning and verification of physics processes, and simulation of neutron and other backgrounds [9]. OSCAR will provide simulated signal and background events for ORCA and will be used to develop and tune parameters and algorithms of the Fast Simulation software sub-system.

WBS 1.3.10 ORCA -- Detector Reconstruction

There is an immediate need to perform detailed studies of the detector and trigger performance such that the design can be optimized. CMS has committed to perform all such studies using OO software. As a result, a high priority "Reconstruction Task Force" was set up in the summer of 1998 under the leadership of D. Stickland of Princeton. This task force had the charge of rapidly building functional OO reconstruction code. The first tangible result of this effort was the successful release of the ORCA (Object-Oriented Reconstruction for CMS Analysis) Version 1 in December 1998. Since then the project has continued to grow and improve the functionality of ORCA [10].

The ORCA task covers software for the reconstruction of simulated (and ultimately real) events. In the case of simulated events, it also includes the simulation of the detector and electronics response to the GEANT hits, the creation of digitized hits and L1 trigger primitives, and the simulation of the L1 trigger response. For both simulated and real data, it provides software for the reconstruction of tracks segments, clusters, etc. within individual subdetectors and certain entities which combine information from more than one subdetector, such as tracks spanning the whole of the tracker.

WBS 1.3.11 OBSELETE

This WBS item previously covered the task of "Physics Object Reconstruction and Higher Level Triggers". Since the task has Software, Physics, and Higher Level Trigger aspects it was decided to set up distinct but closely-linked "Physics Reconstruction and Selection" (PRS) activity. For backwards compatibility, this WBS number is retained but it has no associated tasks or resources.

The PRS task covers software to create physics classes describing reconstructed particle candidates, perform global analyses of complete events, and develop Level 2 and Level 3 trigger algorithms using both localized subsets of data and full event information. The L2 and L3 selection algorithms then need to be evaluated in terms of background rejection power and signal efficiency, taking into account the CPU and bandwidth available in the L2/L3 trigger systems. Four subtasks cover the reconstruction of: electrons and photons; muons; jets and missing energy; and b's and taus.

WBS 1.3.12 IGUANA -- Interactive Graphical User Analysis

This item covers software infrastructure and tools for performing user analysis of real and simulated CMS data [11]. In particular, it covers: user interfaces (UI's); graphical user interfaces (GUI's); statistical and numerical analysis; interactive data analysis and presentation including histogramming; and interactive detector and event visualisation.

For some of these items, toolkits already exist either within HEP or the commercial sector. Therefore, a significant component of this task will be the assessment of available tools and their adaptation, extension, deployment, and support in the CMS environment. In some case, new tools may need to be developed although pre-existing components will be exploited wherever possible. A crucial aspect of this task, which is closely linked with the CARF development, is the provision of user interfaces and the corresponding graphical user interfaces. These will be needed not only for final physics analysis software but also for use by applications based on other CMS software sub-systems, such as ORCA and OSCAR.

WBS 1.3.13 Test Beam

This item covers generic software required for the acquisition, storage, processing, and analysis of test beam data[12]. An important aspect of this work is the deployment and evaluation of the ODBMS and mass storage systems for the storage of test-beam data in a real-time environment.

WBS 1.3.14 Calibration

This item covers software associated to the calibration of the CMS sub-detectors including alignments, energy calibrations, etc. using data from: test beams; laboratory measurements; survey data; in-situ dedicated systems; slow control systems; and physics event samples. The Calibration clients include ORCA and potentially the OSCAR and Fast Simulation projects, which may ultimately incorporate realistic detector performance data into the simulations.

WBS 1.3.15 ODBMS -- Object Database Management System

This item covers generic issues pertaining to the use of an ODBMS for the storage of CMS non-event and event data, including those associated to the operation of a database federation in a heterogeneous distributed environment [13]. In addition, a construction database has been developed and deployed as part of the CRISTAL project.

Objects are stored persistently, reusing as much as possible what is provided by the ODBMS itself to store, organize and retrieve data, and to administer both data and metadata including schema declaration, schema evolution, object versioning and database replication. Optimization mechanisms such as caching, compression, and physical clustering of objects will be implemented such that they are transparent to the user. A large fraction of the ODBMS activities are carried out in the context of the CRISTAL, GIOD, MONARC, RD45, and WISDOM projects.

REFERENCES FOR APPENDIX

1 J-P.Wellisch, *The Status of Software Process Improvement in CMS*. **CMS IN/1999-033**, http://cmsdoc.cern.ch/cms/software/reviews/papers99/SoftwareProcess.ps

2 **ISO/IEC DTR 15504** *Software Process Assessment*, the ISO/IEC Joint Technical Committee 1 (JTC1), Sub-committee 7 (SC7) Working Group 10 (WG10) Project Editor Terry Rout

3 J-P.Wellisch, **CMS NOTE 1998/071**

4 J-P.Wellisch, **CMS NOTE 1998/070**

5 J-P.Wellisch, **CMS NOTE 1998/072**

6 V.Innocente, *CMS Software Architecture: Software framework, services and persistency in high level trigger, reconstruction and analysis*. **CMS IN/1999-034**, http://cmsdoc.cern.ch/cms/software/reviews/papers99/cmsArchitecture1099.ps

7 S.Banerjee, *A Multi-representations Detector Description for CMS Offline Physics Applications***. CMS IN/1999-038**, http://cmsdoc.cern.ch/cms/software/reviews/papers99/cms_detector.ps

8 S.Wynhoff, *Fast Monte-Carlo Simulation in CMS*. **CMS IN/1999-037**. http://cmsdoc.cern.ch/cms/software/reviews/papers99/famos.ps

9 M. Schröder, *CMS Detector Simulation Project OSCAR*. **CMS IN/1999-036**. http://cmsdoc.cern.ch/cms/software/reviews/papers99/oscar.ps

10 D.Stickland, *CMS Reconstruction Software: The ORCA project*. **CMS IN/1999-035**. http://cmsdoc.cern.ch/cms/software/reviews/papers99/cmsORCA.ps

11 G.Alverson, I. Gaponenko and L.Taylor, *IGUANA - Interactive Graphical User Analysis***. CMS IN/1999-042**. http://cmsdoc.cern.ch/cms/software/reviews/papers99/iguana.pdf

12 L.Silvestris, *CMS Test Beam Software*. **CMS IN/1999-043**. http://cmsdoc.cern.ch/cms/software/reviews/papers99/TestBeam.ps

13 I.Willers, *DataBase Activities in CMS*. CMS IN/1999-039**.** **http://cmsdoc.cern.ch/cms/software/reviews/papers99/Database_activities.ps**